JOURNAL ARTICLE

What happens to large changes? Saltation produces well-liked outputs that are hard to generate

Amy Smolek and Vsevolod Kapatsinski

Department of Linguistics, University of Oregon, Eugene, OR, US
Corresponding author: Vsevolod Kapatsinski (vkapatsi@uoregon.edu)

Saltatory alternations 'skip over' intermediate sounds, as in k~s skipping over [t]. Recent research has argued that saltation is diachronically unstable and documented one possible cause of instability: Learners exposed to saltatory alternations may overgeneralize them to intermediate sounds. However, this research has trained participants to criterion or excluded participants who did not reach criterion accuracy on familiar sounds. In first language acquisition, learners of languages with saltatory patterns cannot hope to receive more exposure to the pattern than those learning non-saltatory patterns. For this reason, we examined learning of saltatory and non-saltatory patterns after a constant amount of training. We compared saltatory labial palatalization to non-saltatory alveolar and velar palatalization. Participants showed overgeneralization of saltatory palatalization in a judgment task. However, saltatory alternations did not result in increased rates of palatalizing similar sounds, compared to non-saltatory alternations. Instead, saltatory alternations were less likely to be produced than non-saltatory alternations. These results suggest that large, saltatory alternations may be diachronically unstable because they are harder to (learn to) produce. Instead of being overgeneralized to intermediate sounds, saltatory alternations may disappear from the language by losing productivity and being replaced with faithful mappings.

1. Introduction

Several recent studies have argued that phonological learning is biased against alternations involving dissimilar sounds (Hayes & White, 2015; Moreton & Pater, 2012a, 2012b; Peperkamp, Le Calvez, Nadal, & Dupoux, 2006; Steriade, 2001/2009; White, 2013, 2014, 2017; White & Sundara, 2014; Zuraw, 2007). While dissimilarity between sounds can in principle be defined in terms of phonological features, articulatory gestures, or perceptual dimensions, it is often operationalized in terms of the alternation 'skipping over' an intermediate sound (Peperkamp et al., 2006; Skoruppa, Lambrechts, & Peperkamp, 2011; White, 2013, 2014). For example, the [k]/[s] alternation in *electric/electricity* skips over the intermediate sound [t]. Because of the 'skipping over' metaphor, such alternations have been called 'saltatory' in the literature, and we retain this terminology in the present paper.

The bias against saltatory alternations is thought to give rise to a strong typological tendency: If a language's sound inventory contains three sounds, X, Y, and Z, such that Y falls in between X and Z in similarity space, then X→Z tends to imply Y→Z but not vice versa. Across a number of experimental paradigms, with both children and adults, White and colleagues have documented that exposure to X→Z leads participants to infer that Y would also change into Z in the same context: If even X changes into Z in this context,

then surely Y must as well. Conversely, exposure to Y→Z does not lead participants to infer X→Z. For example, White (2013, 2014) reports that English-speaking adults exposed to p→v prefer f→v to f→f in a forced choice task, whereas those exposed to b→v prefer f→f over f→v.

While these results are consistent with the synchronic rarity of saltatory alternations, their diachronic implications are rather counter-intuitive. If the observed overgeneralization is the mechanism responsible for the implicational universal, then the rarity of saltatory patterns must be ascribed to speakers of languages with saltatory alternations overgeneralizing them to intermediate sounds. Yet, Bybee (2008) has argued that alternations invariably *lose* productivity as their magnitude (i.e., the number of articulatory changes they necessitate) increases. If this is true, then saltatory alternations may be unstable because their large magnitude makes them unproductive and likely to be replaced by faithful mappings. Indeed, the only diachronic study of saltation loss cited in White (2013) documents a saltatory alternation in the midst of losing productivity rather than one being extended to intermediate sounds (Crosswhite, 2000; cited in White, 2013, p. 32). More generally, errors produced by children in generating novel forms of known words largely underapply stem changes rather than extending them to new sounds (Do, 2013; Kerkhoff, 2007; Krajewski, Theakston, Lieven, & Tomasello, 2011; Tomas, van de Vijver, Demuth, & Petocz, 2017). Changes involving overgeneralization of an alternation to an intermediate sound are, in contrast, quite rare (Bolognesi, 1998; White, 2017). Insofar as error can seed language change (see Andersen, 1973; Bybee, 2010; Bybee & Slobin, 1982; Harmon & Kapatsinski, 2017; Hudson Kam & Newport, 2009; for discussion), saltatory alternations may be more likely to disappear via underapplication rather than over-extension of the alternation. In this paper, we attempt to examine the possible fate of large saltatory alternations by testing whether learners tend to underapply or to over-extend them, or both.

Previous experiments have suggested both underapplication and over-extension but did not pit the two possibilities against each other. Skoruppa et al. (2011) found that saltatory alternations like p~z and p~s were harder to learn to produce than non-saltatory alternations like p~t, the former two groups showing slower improvements over the course of the experiment. While Skoruppa et al. evaluated only accuracy on the trained segments, White (2013, pp. 137–146) examined novel segments and found increased overgeneralization of saltatory alternations to intermediate segments (e.g., p→v but not b→v generalizing to f→v). This result contradicts the hypothesis that large alternations are likely to lose productivity. However, White (2013) excluded all participants whose accuracy on trained segments was below 80% from the analysis. This makes it unclear whether accuracy on trained segments would be lower in the saltatory condition if all participants were included.

In the present work, we compare the effects of alternation magnitude on both segments participants were trained to change and those they were not trained to change in both production and judgment tasks within participants. We hypothesized that we would replicate the tendency to extend saltatory alternations to intermediate sounds in the judgment task but would find that participants have difficulties producing large alternations, resulting in their eventual diachronic demise.

1.1. The test case: Palatalization

Like other experiments reviewed above, our research examines the learning of miniature artificial languages differing in the magnitude of base changes they require. In particular, we examined how easy it is to learn to palatalize voiced and voiceless labial, alveolar, or velar stops before -*i*. All experimental conditions had the same output forms (namely,

Smolek and Kapatsinski: What happens to large changes? Saltation
produces well-liked outputs that are hard to generate

Art. 10, page 3 of 27

plurals that ended in -tʃi), but differed in the magnitude of the change required to get from the singular to the plural. Yun (2006) showed that English alveopalatals involve both a tongue body gesture, similar to that of a velar, and a tongue blade gesture, similar to that of an alveolar. In contrast, labials do not share any oral gestures with alveopalatals. Therefore, in articulatory terms, labial palatalization is a larger change than alveolar or velar palatalization and involves saltation over either [t] or [k], depending on whether one reaches [dorsal & coronal] from [labial] through [dorsal] or [coronal].

Typologically, palatalization of coronals and velars is about equally common, while palatalization of labials is extremely rare (Bateman, 2007; Bhat, 1978; Chen, 1973; Kochetov, 2011). Kochetov (2011) proposes an implicational universal: Labial palatalization implies palatalization of coronals and/or velars, while Bateman (2007) suggests that there are *no* clear cases of productive labial palatalization in which the labial articulation is fully suppressed in the palatalized outcome. Typological frequency often maps onto how easy a pattern is to learn (e.g., Finley, 2008; Mitrović, 2012; Schane, Tranel, & Lane, 1975; White, 2013, 2014; White & Sundara, 2014; Wilson, 2006). The typological data thus strongly suggest that there may be a learning bias against labial palatalization, which may cause labial palatalization to either not be learned or to be overgeneralized to coronals and/or velars, resulting in patterns obeying the implicational universal.

1.2. Preliminary work

The present work follows up on preliminary results previously reported in a proceedings paper (Stave, Smolek & Kapatsinski, 2013). In this previous work, we also exposed participants to one of three different palatalization patterns: p→tʃ, k→tʃ, or t→tʃ, but palatalization was triggered by -*a*, and did not occur before -*i*. This was to show that an observed bias against changing some consonants could not be attributed to high markedness of unchanged consonants in the change-triggering context. We found that the differences in palatalization rates between the to-be-palatalized consonant and the not-to-be-palatalized consonants were significantly smaller in the group trained on p→tʃ than in the other groups, despite the fact that there is nothing particularly marked about [pa]. The interaction appeared to be driven both by a lower palatalization rate for to-be-palatalized labials compared to to-be-palatalized alveolars and velars and by increased palatalization rates for not-to-be-palatalized consonants in the labial condition. However, neither of the individual effects reached significance. Judgment data showed a trend towards the same interaction but it was not statistically significant.

In the present study, we replicate Stave et al. (2013) using palatalization before [i], rather than [a], a more natural context that we hoped would make palatalization easier to learn (Mitrović, 2012; Wilson, 2006). In this context, palatalization of [t] and [k] is also plausibly driven by the markedness of [ti] and [ki]. Examining palatalization before -*i* therefore lets us test whether the bias against labial palatalization observed cross-linguistically in natural contexts (Bateman, 2007; Kochetov, 2011) is driven by the low markedness of [pi] compared to [ti] and [ki]. We also changed the distribution of the palatalizing and non-palatalizing suffix across consonants, with the aim of increasing the experiment's power to detect differences in palatalization rates across conditions. The languages in Stave et al. (2013) used the palatalizing suffix 50% of the time each with the to-be-palatalized stems, and 100% of the time elsewhere. The present languages reverse these proportions, so that the palatalizing suffix is favored rather than disfavored by to-be-palatalized consonants. The aim of this change was to increase the number of instances of palatalization in the appropriate context in two ways. First, we hoped for an increase in the use of the palatalizing suffix with to-be-palatalized consonants: Participants tend to match suffix probabilities, at least in the aggregate (e.g., Kapatsinski, 2010). Second, Kapatsinski

(2010) showed that exposing participants to many examples of a suffix simply attaching to a stem without changing it reduces the productivity of changes triggered by the suffix. For example, experiencing many examples of t→ti and p→pi reduces the productivity of k→tʃi relative to k→ki. We hoped to maximize the productivity of palatalization by ensuring that the palatalizing suffix changes the stem it attaches to more often than not.

1.3. The languages

In our experiment, participants were placed in either Labial, Alveolar, or Velar Palatalization conditions, which contained the corresponding palatalization pattern, as illustrated in **Table 1**. Singulars were always C(C)VC forms ending in oral stops. Plurals were formed by adding the plural suffixes -*i* and -*a.* Depending on the condition to which a participant was assigned, either labial, alveolar, or velar stops were palatalized, becoming [tʃ] if voiceless and [dʒ] if voiced. (The voiced varieties were included to test for the effect of articulatory vs. perceptual similarity, a question that is outside the scope of the present paper.) We consider labial palatalization to be saltatory because it involves changing a labial into a sound that involves both coronal and dorsal articulations (Yun, 2006).

Participants' knowledge of the language was tested using elicited production and judgment tests requiring generalization to novel singulars. The production test involved hearing a novel singular and saying the corresponding plural form. The judgment test involved hearing singular-plural pairs and judging whether the plural form is the right one for the singular form.

1.4. Hypotheses

As discussed above, our basic expectation was that labial palatalization—a large saltatory alternation—would be harder to learn than alveolar or velar palatalization. In particular, we expected the p~tʃ alternation to be more difficult to produce after training than the k~tʃ and t~tʃ alternations. In addition, if saltatory alternations overgeneralize to intermediate sounds and [tʃ] is both coronal and dorsal while [p] is neither, labial palatalization should be taken to imply either velar or alveolar palatalization but not vice versa. We now describe a number of more specific hypotheses we evaluate. These hypotheses are used to structure the results section.

Hypothesis 1: Labial palatalization is hard to learn because of faithfulness, not markedness. Judgments of unfaithful mappings will differ across conditions but judgments of faithful mappings will not. In other words, p~tʃ is worse than k~tʃ or t~tʃ but k~ki and t~ti are as good as p~pi. This finding must hold for us to argue that the bias against labial palatalization is a bias against certain *alternations* rather than against certain structures

Table 1: Labial, Alveolar, and Velar Palatalization patterns presented to participants in Experiment 1.

Singular	Labial Palatalization Plural	Alveolar Palatalization Plural	Velar Palatalization Plural
…p	**…tʃi**	…{pi;pa}	…{pi;pa}
…b	**…dʒi**	…{bi;ba}	…{bi;ba}
…t	…{ti;ta}	**…tʃi**	…{ti;ta}
…d	…{di;da}	**…dʒi**	…{di;da}
…k	…{ki;ka}	…{ki;ka}	**…tʃi**
…g	…{gi;ga}	…{gi;ga}	**…dʒi**

Smolek and Kapatsinski: What happens to large changes? Saltation
produces well-liked outputs that are hard to generate

Art. 10, page 5 of 27

the alternations repair (e.g., Gnanadesikan, 1997; Kirchner, 1996; Steriade, 2001/2009; White, 2013). If Hypothesis 1 does not hold for our data, any condition differences we obtain may be due to markedness differences between [pi] on the one hand and [ti] and [ki] on the other, the structures repaired by palatalization. In particular, if [ki] and [ti] are worse (i.e., more marked) than [pi], palatalization of [t] and [k] is expected to be easier to learn than palatalization of [p] on these grounds alone (Pater & Tessier, 2006).

Hypothesis 2: Large alternations are hard to produce. If alternation magnitude influences learnability, labial palatalization should be harder to learn than alveolar or velar palatalization. In particular, if large alternations (of which saltation is one example) are hard to learn to produce (Skoruppa et al., 2011), [p] should be palatalized less in the Labial condition than [t] is palatalized in the Alveolar condition or [k] is palatalized in the Velar condition.

Hypothesis 3: Saltatory alternations are likely to be overgeneralized. If saltatory alternations are especially likely to be overgeneralized (Hayes & White, 2015; Moreton & Pater, 2012a; White, 2013, 2014, 2017; White & Sundara, 2014), the same not-to-be-palatalized consonants should be palatalized more in the Labial condition than in the other conditions. For example, [t] might be palatalized more often when participants are trained on p→tʃi than when they are trained on k→tʃi.

Hypothesis 4: Large changes are hard to produce even if judged to be preferable. The difference between the Labial condition and the other conditions should be greater in the production test than in the judgment test. Particularly, we hypothesized that labial palatalization may be rarely produced even when judged to be more acceptable than lack of palatalization. This expectation was based on a small-scale experiment reported by Zuraw (2000). Zuraw found that Tagalog speakers seldom produced nasal substitution, which requires a change to the base, but judged it to be more acceptable than nasal assimilation, which does not require a stem change. This dissociation between judgment and production would also address the possible criticism that judgments are simply more tolerant than production (e.g., Kempen & Harbusch, 2005) by showing that the less acceptable alternative is more likely to be produced. In other words, it would indicate that the speaker recognizes that the form that results from an alternation is the right thing to say without quite having the ability to say it.

2. Methods

2.1. Participants

One hundred seven undergraduate students in introductory psychology or linguistics courses at the University of Oregon were recruited through the Human Subject Pool and received partial course credit for participation. Eleven subjects were excluded for producing plurals that did not correspond to the patterns in the training. This left 32 subjects in the Alveolar condition, 31 in the Labial, and 33 in the Velar. All reported being native speakers of English, with no speech, hearing, language, or learning disabilities.

2.2. Training

In every condition, there were 28 unique singular forms of words, which were randomly paired with Spore creatures selected from the database of user-created images at sporepedia.com. Each creature was shown at least once alone, and at least once as part of a group (the same image copied and pasted multiple times); the images containing a single creature were accompanied by a recording of the singular form, and the images containing groups of creatures were paired with the corresponding plural form (see **Figure 1** below for illustration). All word-picture pairings appeared in random order, so the corresponding singulars and plurals were rarely temporally adjacent. In previous

Figure 1: Example display for the Labial Palatalization condition. Participants saw the creature(s) and heard the associated word (shown in brackets here). The trial order was random.

work (Kapatsinski, 2012), this ordering was found to encourage overgeneralization compared to presenting singular-plural pairs, presumably by making it less obvious *which* singular-final consonants map onto [tʃ] and [dʒ].

The singular forms for all tokens were single-syllable roots of C(C)VC structure, where the word-final consonant was an oral stop. Twelve out of the 28 training words ended with the to-be-palatalized consonants (Alveolar contained 6 each of stems ending with [t] and [d], Labial of [p] and [b], and Velar of [k] and [g]), with the remaining 16 split evenly between the other two place*voice combinations. Stimuli are shown in the appendix.

Note that participants were presented with the same amount of training regardless of whether they were trained on labial, alveolar, or velar palatalization. This contrasts with some previous work, which has trained participants to criterion or excluded low performers (White, 2013, 2014). We believe that these procedures obscure the nature of the bias against large changes by ensuring that participants learn the trained alternations equally well.

2.3. Production test

For the production test portion of the experiment, we selected an additional 92 creatures from the Spore Creature Creator database. As before, we copied and pasted the creatures such that, for each creature, there was one slide showing a single creature and one slide showing a group of creatures. The singular slide was again accompanied by the recording of the singular form, but the plural slide played no recording. Instead, subjects were told to say what they thought the plural form for the singular should be, which we recorded for later coding. Thirty-six of the 92 singulars ended with the consonants subjects had been trained to palatalize (half voiced, half voiceless), with the remaining 56 split evenly between the other places and voicing.

Each plural recording from a subject was saved as a separate file and coded by either the first author or undergraduate RAs, whose codings were all checked by the first author for accuracy. In case of inter-coder disagreement, the first author re-listened to the file and coded it appropriately; in the rare case that the production was unclear, the spectrogram was examined. We coded in particular for the stem-final consonant and the word-final vowel in the plural form. If a participant replaced the stem-final consonant with either of the palatals (e.g., *blaɪp → blaɪtʃi*), we coded the form as being palatalized; if the participant retained the same consonant in the plural (e.g. *blaɪp → blaɪpi*), the form was coded as not palatalized. If the stem consonant was retained but the palatal was also added (e.g., *blaɪp → blaɪptʃi*), we also categorized the form as not palatalized. While these products fit the "plurals should end in tʃi" schema, they preserve the final consonant of the base, underapplying the alternation. Biases to preserve aspects of the base should therefore

Smolek and Kapatsinski: What happens to large changes? Saltation
produces well-liked outputs that are hard to generate

Art. 10, page 7 of 27

favor these kinds of plurals alongside plurals like *blaɪp* → *blaɪpi*. Because these 'blend' *Ctʃi* responses are relatively rare, the results are unchanged if they are omitted or coded as palatalized instead. Rarely, a subject would replace the stem consonant with another non-palatal (e.g. *blaip* → *blaida*); such responses are excluded from all analyses. Instances where subjects added the English plural *-s* were also excluded, and subjects who produced a majority of such responses were excluded entirely.

2.4. Judgment test

Since the judgment task exposes participants to plural forms that contradict training, it followed the production task. The judgment session was the same across all conditions. Thirty new singulars were created, divided equally between the three places and two voices. For each singular, we recorded four plural forms, crossing whether the final consonant was palatalized and whether the suffix was *-i* or *-a* (e.g., for the alveolar stem *prut*, the plural forms were *pruti, pruta, prutʃi,* and *prutʃa*). For each ratings trial, the picture of a single creature appeared first with the recording of the singular form, followed by a picture of a group of the same creatures with the recording of one of the plural forms.

Subjects were instructed to indicate, using a button box, whether the plural form they heard was the right plural form for the singular. Each subject received a different random order of the 120 singular-plural pairings.

The button box had five buttons but the participants' responses were strongly bimodal: 59% of the ratings were either '1' or '5,' 29% were '2' or '4,' and only 12% were '3.' Because of this, and for comparison with the inherently binary production task data, we transformed the obtained ratings into binary dependent variables. See Appendix B for analyses using untransformed data. Two such variables were created:

(1) Absolute Judgment, which was simply the binarized rating, so for every singular-plural pair we entered '1' (accepted) for ratings above 3 or '0' (rejected) for ratings below 3. We excluded the 12% of ratings that were '3,' since we took them as an indication of indecision.

(2) Relative Judgment, which was the binarized difference of ratings between the palatalized and non-palatalized plurals with the same base and the same suffix by a particular subject, e.g., the rating of *bup~bupi* minus the rating of *bup~butʃi*. It was possible for a subject to rate both plurals equally, in which case the trial was excluded.

The Absolute Judgment variable allows us to examine the effects of condition on judgments of palatalization and faithful mappings separately. This is essential to evaluate Hypothesis 1, that the observed bias is driven by faithfulness rather than markedness. Previous studies of the bias against saltation (except for our preliminary data in Stave et al., 2013) have forced participants to choose between a faithful and an unfaithful form by using production or forced choice tasks, which does not allow the experimenter to distinguish between a preference for one form and a dislike of the other (when given the singular *blup* and the plural options *blupa* and *blutʃa*, they may choose *blupa* because they really like it, or because they really dislike *blutʃa*; see Kapatsinski, 2007, for discussion).

The Relative Judgment variable was originally created to compare judgment to production. It embodies the assumption that palatalization is produced when it is more acceptable than lack of palatalization. However, there were no effects of training condition on judgments of faithful mappings. Consequently, the results are the same whether Absolute or Relative Judgments are used, and only the more informative Absolute Judgment data are reported below.

Art. 10, page 8 of 27

Smolek and Kapatsinski: What happens to large changes? Saltation produces well-liked outputs that are hard to generate

While some readers may be concerned that 3's were excluded from the binary judgment analysis, the distribution of 3's is equivalent across training conditions and final consonant place (**Figure 2**; for training conditions: F(2) = 1.33, *p* = 0.26, *ns,* left panel; for final consonant place: F(2) = 0.98, *p* = 0.38, *ns,* right panel), and all results reported below are the same regardless of whether binarized ratings or the full rating scale is used. As such, we report the results of the analyses based on absolute judgments in the text for ease of comparison to production, and the results using the untransformed rating scale in Appendix B.

2.5. Statistical analysis

All statistical analyses were conducted using generalized (logistic) linear mixed-effects models by means of the lme4 package (version 1.1–9, Bates et al., 2015) in R (version 3.1.1, R Development Core Team, 2014). We included fixed effects for Training Condition (contrast coded as noted for each model), Plural Vowel (*-i* vs *-a*), Test Place (Labial, Alveolar, Velar), To-Be-Palatalized Place (To-be-palatalized vs. Not-to-be-palatalized given the training), Test Voice (Voiced vs. Voiceless), Test Type (Production vs. Judgment), and interactions as applicable. Random intercepts were included for Subjects and Bases, and we used the full random effect structure that would allow our models to converge. Log likelihood tests on nested models were used to derive significance values. When a contrast was not significant, and this null result was expected under some hypothesis, evidence for the null hypothesis was evaluated using the BIC approximation to the Bayes Factor (Wagenmakers, 2007). The BIC approximation to the Bayes Factor compares the posterior probabilities of the null and alternative hypotheses under the assumption that their prior probabilities are equal. Unlike a frequentist analysis, it can therefore provide positive evidence in favor of the null hypothesis, distinguishing between 'lack of evidence against the null' and 'evidence for the null.' Tested models are presented in the notes. The full dataset and analysis code are available as additional files below.
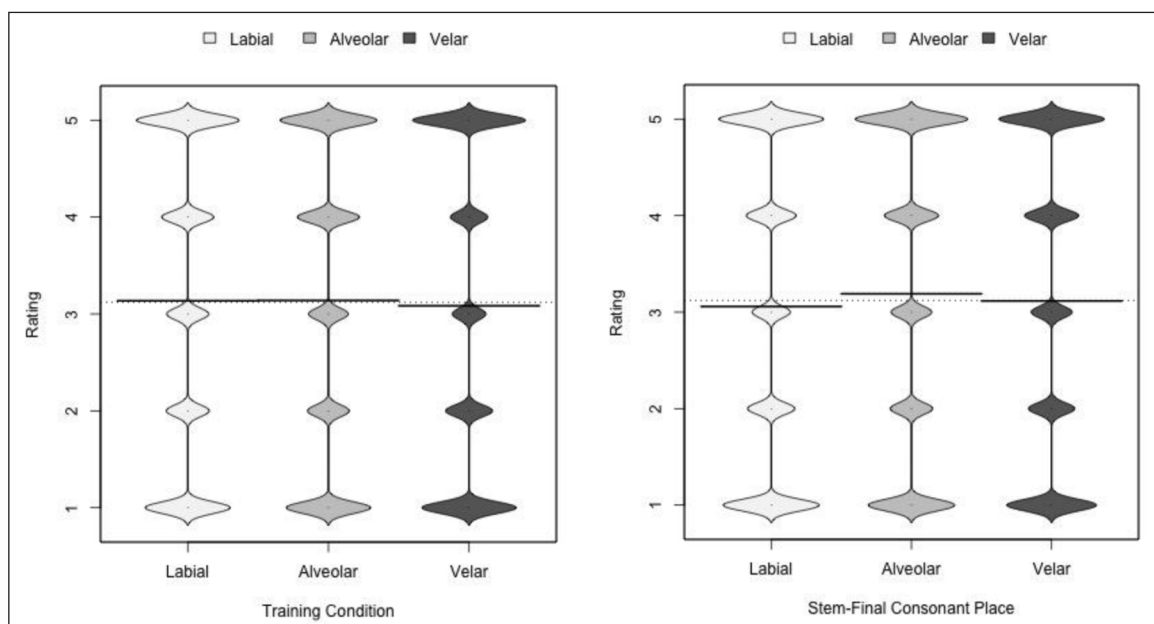

Figure 2: Distribution of ratings by training condition (left) and final consonant place (right) by training condition. The black bars indicate the means by factor level; the dotted line indicates the overall mean.

Smolek and Kapatsinski: What happens to large changes? Saltation produces well-liked outputs that are hard to generate

Art. 10, page 9 of 27

3. Results

3.1. Hypothesis 1: The bias against labial palatalization is not due to markedness

Before we turn to determining whether there is a bias against p~tʃ, we would like to evaluate whether any such bias could be explained as a side effect of the markedness of [ti] and [ki] compared to [pi]. Perhaps [t] and [k] are easy to change into [tʃ] compared to [p] before [i] because participants come to the experiment ready to avoid [ti] and [ki] but have no pre-existing aversion to [pi]. We can evaluate this possibility using the absolute judgments of faithful mappings in the judgment test. If there is a bias against [ti] and/or [ki], we would expect judgments of p~pi in the Labial condition to be higher than judgments of t~ti in the Alveolar condition or k~ki in the Velar condition. However, **Figure 3** shows no significant difference in ratings of no-change plurals for to-be-palatalized consonants across conditions[1] and, if anything, a slight trend in the unexpected direction for not-to-be-palatalized consonants,[2] which is also not significant (**Tables 2** and **3**). According to the BIC approximation to the Bayes Factor, these results provide very strong evidence for the null hypothesis (ΔBIC = 14.1, P_{BIC} (H_0|D) = 0.999). These results are inconsistent with the markedness explanation: The bias is against certain changes, not against certain output structures. In addition, they show that subsequent analyses can continue to utilize absolute rather than relative judgments as the dependent variable: Any between-condition differences in relative judgments would be driven by differences in absolute judgments of alternations rather than of faithful mappings.
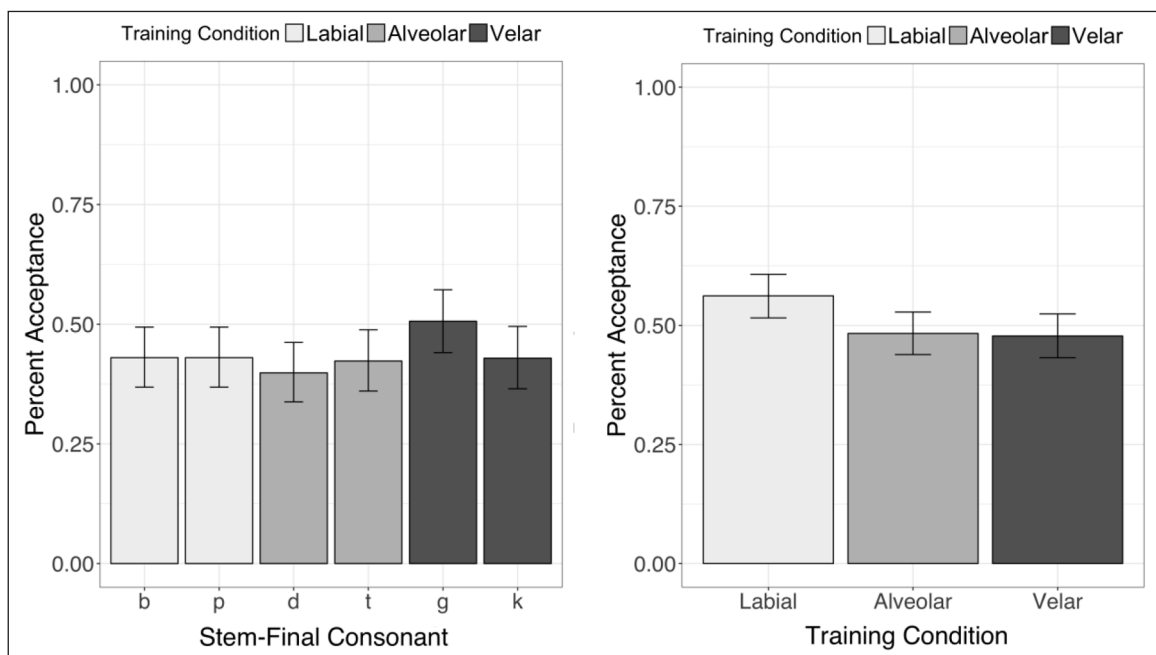
Figure 3: Judgments of faithful plurals do not significantly differ across conditions. Left panel: Acceptance of incorrect faithful plurals, across suffixes. Right panel: Acceptance of correct faithful plurals, before -*i*.

---

[1] Rating Bin ~ Training Condition + Test Voice + Plural Vowel + (1 + Test Voice + Plural Vowel|Subject) + (1 + Training Condition + Plural Vowel|Base); Helmert contrast coded for Labial vs. Alveolar and Velar training, and Alveolar vs. Velar training.

[2] RatingBin ~ Training Condition + Test Voice + (1 + Test Voice|Subject) + (1 + Training Condition|Base); Helmert contrast coded for Labial vs. Alveolar and Velar training, and Alveolar vs. Velar training.

Table 2: Judgments of incorrect faithful mappings for to-be-palatalized consonants across training conditions. The inclusion of Training Condition does not significantly improve the fit of the model, $\chi^2(2) = 0.45$, $p = 0.80$, *ns*.

	b	*se(b)*	*z*	*p*
(Intercept)	−0.48295	0.22952	−2.104	0.0354*
Labial vs. Alveolar and Velar Training	−0.08744	0.36176	−0.242	0.809
Alveolar vs. Velar Training	−0.27456	0.42757	−0.642	0.5208
Voiceless	−0.05531	0.15829	−0.349	0.7268
-a	0.35555	0.20704	1.717	0.0859.

Table 3: Judgments of correct faithful mappings for not-to-be-palatalized consonants across training conditions; Training Condition does not significantly improve the fit of the model, $\chi^2(2) = 0.098$, $p = 0.95$, *ns*.

	b	*se(b)*	*z*	*p*
(Intercept)	0.29886	0.13278	2.251	0.0244*
Labial vs. Alveolar and Velar Training	−0.07223	0.24345	−0.297	0.7667
Alveolar vs. Velar Training	−0.03532	0.28153	−0.126	0.9002
Voiceless	0.1244	0.10086	1.233	0.2174
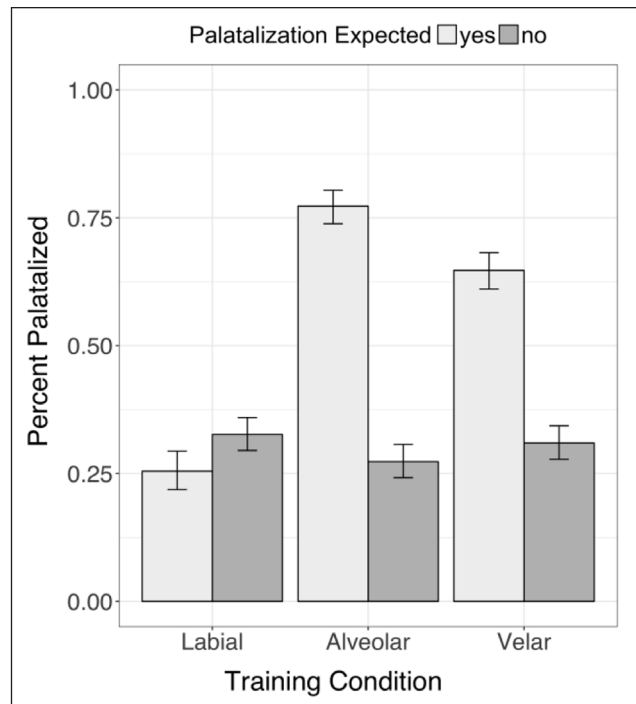
Figure 4: Correct vs. incorrect palatalization rates before -*i* in production by training condition.

3.2. Hypothesis 2: Large alternations are hard to produce

Each bar in **Figure 4** represents the rate of palatalization in production for the to-be-palatalized consonants (light bars) and not-to-be-palatalized consonants (dark bars). Bars are grouped by training language, which determines the identity of the to-be-palatalized consonants. For example, after exposure to the Labial language, the to-be-palatalized consonants are labial. According to Hypothesis 2, to-be-palatalized consonants should be palatalized less often

Smolek and Kapatsinski: What happens to large changes? Saltation produces well-liked outputs that are hard to generate

Art. 10, page 11 of 27

when the palatalization would require a large change to the base, i.e., in the Labial condition. **Figure 4** shows that there is a large difference between the Labial condition and the lingual conditions in the expected direction: To-be-palatalized labials are palatalized dramatically less often than to-be-palatalized alveolars or velars. Participants learned to palatalize velars when trained on velar palatalization, and alveolars when trained on alveolar palatalization. However, they did not learn to palatalize labials any more than non-labials when exposed to labial palatalization. In fact, participants exposed to labial palatalization palatalized labials slightly less than non-labials. These differences are largely due to how often the to-be-palatalized consonants are palatalized (light bars), as expected under Hypothesis 2, rather than in how much palatalization is overgeneralized to the not-to-be-palatalized consonants (Hypothesis 3). These results replicate Skoruppa et al. (2011) and contradict White (2013).

Statistically, we observe that to-be-palatalized base consonants are palatalized significantly less often (i.e., retained significantly more often) in the Labial Palatalization condition relative to the Alveolar and Velar Palatalization conditions (**Table 4**),[3] with no overall effect of voicing and no interactions with voicing. Training Condition significantly improved the fit of the model, $\chi^2(2) = 29.47$, $p < 0.001$.

Hypothesis 2, a bias against changing labials into alveopalatals, predicts that labials will be palatalized less than the linguals both when they should be palatalized, and when they should *not* be palatalized. The first half of the claim is confirmed by **Figure 4** and **Table 4**. The second half of the claim is confirmed by the data in **Figure 5** and **Tables 5**[5] and **6**[6]: Labials are palatalized in error less than other stops are palatalized in error; overgeneralization to labials is shown in dark bars and overgeneralization to linguals in light bars.

To summarize the results so far, the data strongly support a bias against labial palatalization in production. There is a lower rate of palatalization for both the (to-be-palatalized) labials in the Labial training condition and the (not-to-be-palatalized) labials in the other training conditions. In other words, participants are less likely to palatalize labials than linguals, whether correctly or in error.

3.3. Hypothesis 3: Saltatory alternations are likely to be overgeneralized

In order to evaluate this hypothesis, we ask whether alveolars are palatalized more when participants are trained to palatalize labials than when they are trained to palatalize velars. Similarly, are velars palatalized more when participants are trained to palatalize

Table 4: The effect of Training Condition on (erroneous) retention rates of to-be-palatalized consonants in production, before *-i*. Negative regression coefficients indicate higher rates of palatalization (less retention of the base consonant), which in this case means higher accuracy.
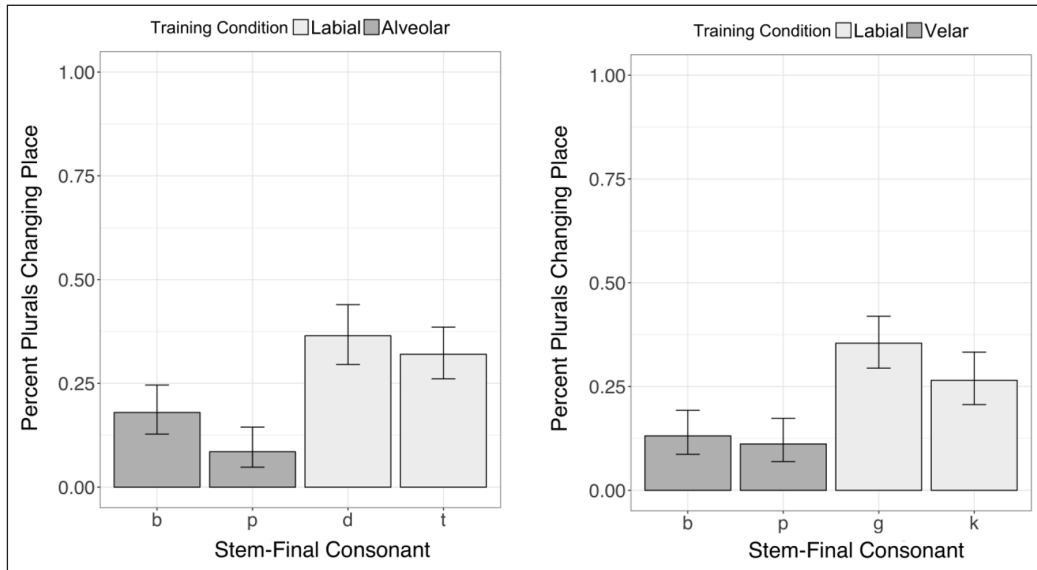
	b	se(b)	z	p
(Intercept)	0.01121	0.36582	0.031	.976
Labial vs. Alveolar and Velar Training	4.17856	0.80836	5.169	<.00001***
Alveolar vs. Velar Training	−1.20023	0.86301	−1.391	.164[4]
Voiceless	0.12622	0.26445	0.477	.633

---

[3] Keep Place ~ Training Condition + Test Voice + (1 + Test Voice|Subject) + (1|Base); Helmert coded for Labial vs. Alveolar and Velar, and Alveolar vs. Velar.

[4] According to the BIC approximation to the Bayes Factor, the results provide positive evidence for the null (ΔBIC = 3.8, P_{BIC} (H0|D) = 0.87).

[5] Keep Place ~ Test Voice + Test Place + (1 + Test Voice|Subject) + (1|Base); restricted to Alveolar training palatalization of labials and Labial training palatalization of alveolars.

[6] Keep Place ~ Test Voice + Test Place + (1 + Test Voice|Subject) + (1|Base); restricted to velar training palatalization of labials and labial training palatalization of velars.

Art. 10, page 12 of 27

Smolek and Kapatsinski: What happens to large changes? Saltation produces well-liked outputs that are hard to generate

Figure 5: Production probabilities for incorrect palatalization as a function of stem-final consonant and training condition. Left panel: Overgeneralization of alveolar palatalization to labials (light) vs. over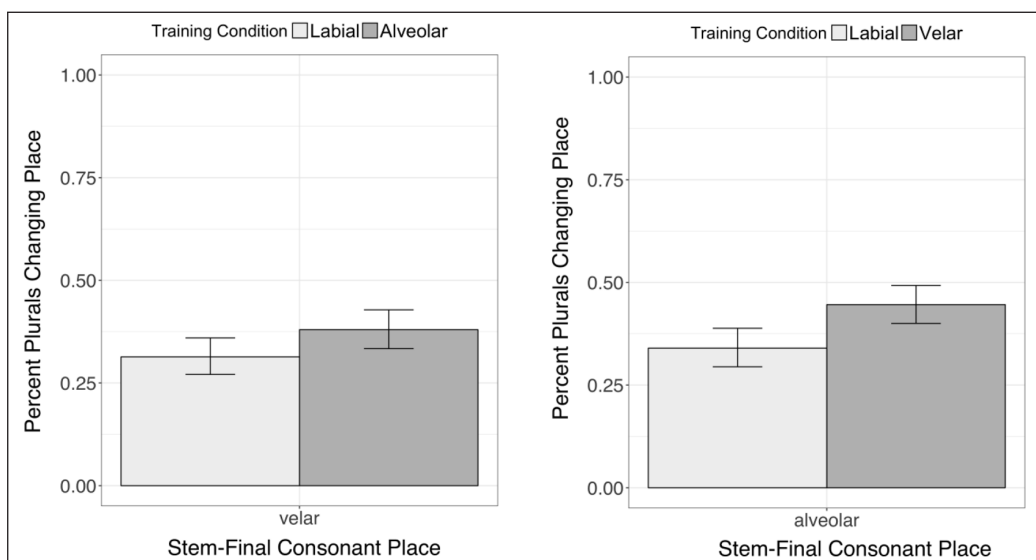generalization of labial palatalization to alveolars (dark). Right panel: Overgeneralization of velar palatalization to labials (light) vs. overgeneralization of labial palatalization to velars (dark).

Table 5: Alveolars are palatalized after training on labial palatalization more than labials are palatalized after training on alveolar palatalization. Test Place significantly improves the fit of the model, $\chi^2(1) = 13.13$, $p < 0.001$.

	b	se(b)	z	p
(Intercept)	4.2555	0.9343	4.555	<.00001***
Voiceless	0.7503	0.7457	1.006	.31432
Alveolar Stem	−3.3044	1.0073	−3.28	.00104**

Table 6: Velars are palatalized after training on labial palatalization more than labials are palatalized after training on velar palatalization, with Test Place significantly improving the model fit, $\chi^2(1) = 7.13$, $p = 0.008$.

	b	se(b)	z	p
(Intercept)	3.8124	0.784	4.863	<.00001***
Voiceless	0.9275	0.7749	1.197	.23131
Velar Stem	−2.3129	0.8975	−2.577	.00996**

labials than when they are trained to palatalize alveolars? If alveopalatals are [dorsal] and [coronal] (Yun, 2006), then changing a labial into an alveopalatal involves changing both features and therefore saltation over either [t] or [k], depending on the route taken. In contrast, a change from the [dorsal]-only [k] to [dorsal;coronal] does not involve saltation over a [coronal]-only sound, [t]. Similarly, a change from [coronal]-only [t] to [coronal;dorsal] does not involve saltation over [k]. Thus, labial palatalization should overgeneralize to velars and alveolars but velar and alveolar palatalization need not support each other.

Smolek and Kapatsinski: What happens to large changes? Saltation produces well-liked outputs that are hard to generate

Art. 10, page 13 of 27

Figure 6: Overgeneralization of palatalization in production depending on magnitude. Left panel: Overgeneralization of palatalization to velar-final stems after Labial Palatalization (light bar) vs. Alveolar Palatalization (dark bar) training. Right panel: Overgeneralization of palatalization to alveolar-final stems after Labial Palatalization (light bar) vs. Velar Palatalization (dark bar) training.

Contrary to Hypothesis 3, there was no significant difference in palatalization rates of alveolars between the Labial and Velar training conditions[7] ($b = -0.46$, $se(b) = 0.68$, $z = -0.67$, $p = 0.50$). Likewise, there was no significant difference in palatalization rates of velars between the Labial and Alveolar training conditions[8] ($b = -0.46$, $se(b) = 0.67$, $z = -0.69$, $p = 0.49$); see **Figure 6**. Furthermore, based on the BIC approximation to the Bayes Factor, the results provide strong evidence for the null hypothesis in both cases: $\Delta BIC = 6.9$, $P_{BIC}(H_0|D) = 0.97$ and $\Delta BIC = 7$, $P_{BIC}(H_0|D) = 0.97$, respectively. In other words, the present study had sufficient power to provide positive evidence in favor of the null hypothesis that overgeneralization is no more likely with a saltatory change than with a non-saltatory change (contra White, 2013, 2014).

In the judgment data, subjects trained on Labial and Alveolar Palatalization did not differ in their judgments of k→tʃ overall[9] ($b = -0.02$, $se(b) = 0.31$, $z = -0.70$, $p = 0.94$, *ns*) or in the palatalizing context, before *-i*[10] ($b = -0.40$, $se(b) = 0.44$, $z = -0.91$, $p = 0.36$, *ns*); based on the BIC approximation to the Bayes Factor, there is strong evidence for the null across suffixes ($\Delta BIC = 6.8$, $P_{BIC}(H_0|D) = 0.97$) and positive evidence before *-i* ($\Delta BIC = 5.4$, $P_{BIC}(H_0|D) = 0.94$). **Figure 7** shows the acceptability of k→tʃ by training language across suffixes and before *-i*.

Velar and Labial training conditions did not show a difference in judgments of t→tʃ across suffixes[11] ($b = -0.63$, $se(b) = 0.39$, $z = -1.59$, $p = 0.11$, *ns*. According to the BIC approximation to the Bayes Factor, the results are positive evidence for the null hypothesis

---

[7] Keep Place ~ Training Condition + Test Voice + Plural Vowel + (1 + Test Voice + Plural Vowel|Subject) + (1 + Training Condition + Plural Vowel|Base); restricted to velar and labial training palatalization of alveolar stems.

[8] Keep Place ~ Training Condition + Test Voice + Plural Vowel + (1 + Test Voice + Plural Vowel|Subject) + (1 + Training Condition + Plural Vowel|Base); restricted to alveolar and labial training palatalization of velar stems.

[9] RatingBin ~ Training Condition + Test Voice + Plural Vowel + (1 + Test Voice + Plural Vowel|Subject) + (1 + Training Condition|Base), restricted to labial and alveolar training ratings of palatalized velars.

[10] RatingBin ~ Training Condition + Test Voice + (1 + Test Voice|Subject) + (1 + Training Condition|Base), restricted to labial and alveolar training ratings of palatalized velars before *-i*.

[11] RatingBin ~ Training Condition + Test Voice + Plural Vowel + (1 + Test Voice + Plural Vowel|Subject) + (1 + Training Condition + Plural Vowel|Base); restricted to labial and velar training ratings of palatalized alveolars.
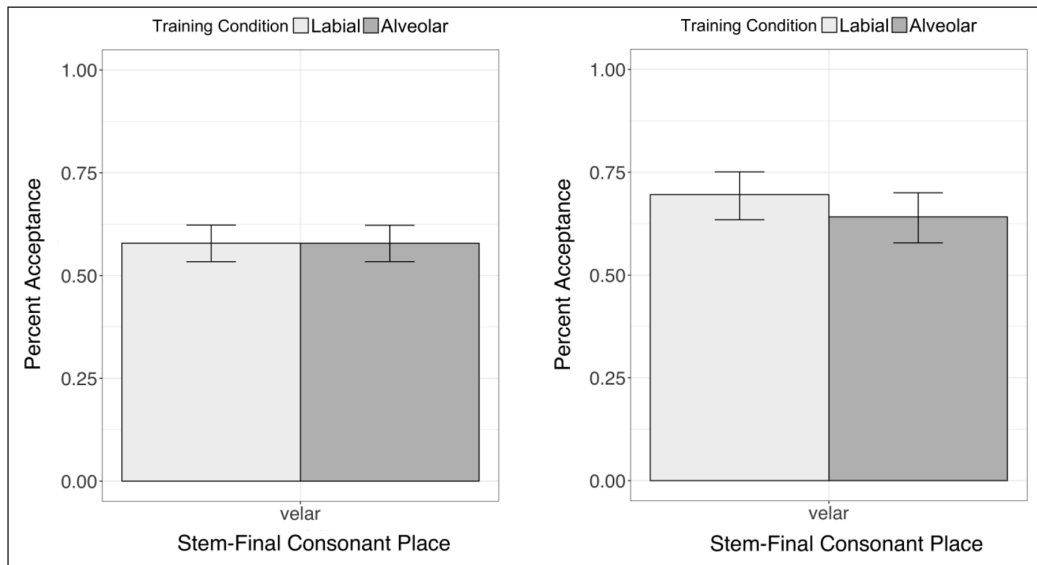
Figure 7: Acceptance of overgeneralization of palatalization to velars in judgment after Labial Palatalization (light bars) vs. Alveolar Palatalization (dark bars) training. Left panel: Across suffixes. Right panel: Before *-i*.

(ΔBIC = 4.4, P_{BIC} (H_0|D) = 0.90). In the palatalizing context, subjects trained on velar palatalization gave 'marginally' lower ratings of palatalized alveolars than those trained on labial palatalization (b = –0.89, *se(b)* = 0.47, z = –1.86, p = 0.06; Trained Place marginally improves model fit, $\chi^2(1)$ = 3.40, p = 0.06). This trend is in the direction consistent with greater overgeneralization in the Labial condition. However, according to the BIC approximation to the Bayes factor, the results are still more consistent with the null than with the alternative hypothesis, ΔBIC = 2.78, P_{BIC} (H_0|D) = 0.80). **Figure 8** shows the acceptance rates of t→tʃ after Labial and Velar training.
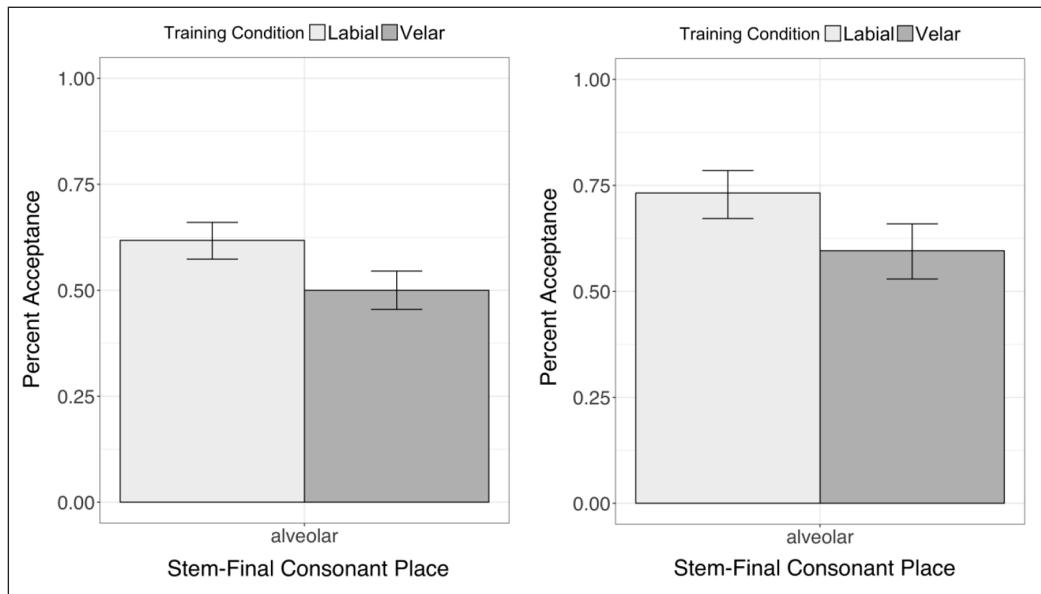
Overall, there is little evidence that saltatory alternations are especially likely to be overgeneralized, even in a judgment task. In any case, the difference in likelihoods of overgeneralization between saltatory and non-saltatory alternations is quite small, compared to the very large differences in productivity reported in **Figure 4**.

3.4. Hypothesis 4: Large changes are hard to produce even if judged to be preferable

We suspected that the bias against labial palatalization should be stronger in production than in judgment. Since the judge is presented with the palatalized output form, whatever difficulty they would face generating the form is alleviated (see also Harmon & Kapatsinski, 2017; Luce & Pisoni, 1998, for similar arguments regarding the difference between open-set and closed-set tasks).

As we saw in **Figure 2** above, ratings of faithful plurals do not differ across conditions. Thus, any judgment differences between conditions we could see must come from judgments of unfaithful forms featuring palatalization. These data are shown in the left panel of **Figure 9**, side by side with production data from **Figure 4** repeated here in the right panel.

Table 7 reports a mixed-effects regression model comparing the patterns of results across the two panels of **Figure 9**. The dependent variable for the model in **Table 7** was binary, with 0 corresponding to production of a non-palatalized form in the production task and a rejection (rating < 3) of a palatalized form in the judgment task; 1 corresponded to production of a palatalized plural in the production task and acceptance (rating > 3) in the judgment task. The comparison reveals several findings. First, there is a striking

Smolek and Kapatsinski: What happens to large changes? Saltation produces well-liked outputs that are hard to generate

Art. 10, page 15 of 27

Figure 8: Acceptance of overgeneralization of palatalization to alveolars in judgment after Labial Palatalization (light bars) vs. Velar Palatalization (dark bars) training. Left panel: Across suffixes. Right panel: Before *-i*.
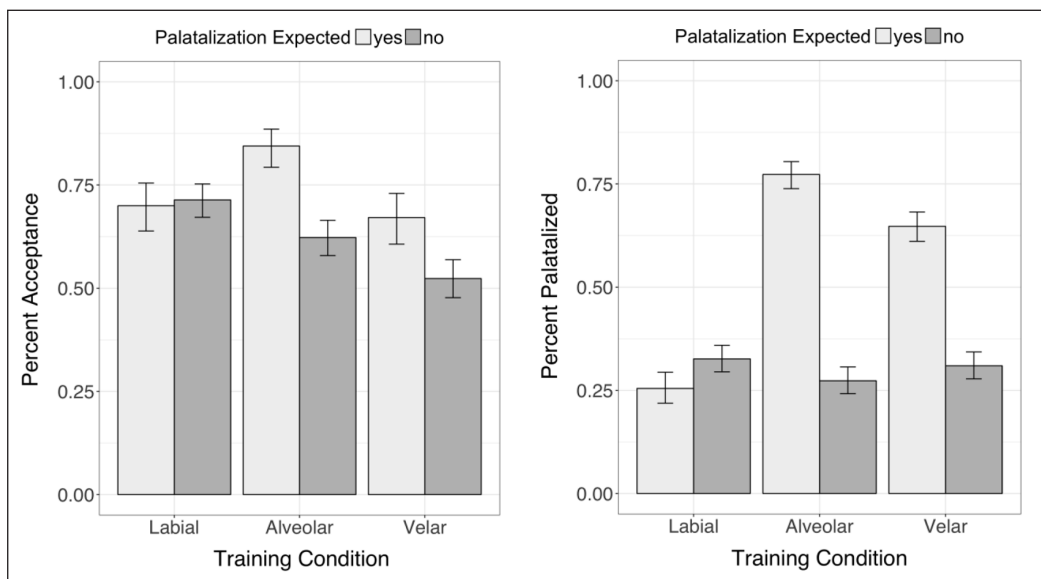
Figure 9: Correct (light bars) vs. incorrect (dark bars) palatalization by training condition, before *-i*. Left panel: Acceptance of labial palatalization in perception. Right panel: Lack of labial palatalization in production.

Table 7: The effects of training on Labial vs. Alveolar and Velar Palatalization on judgment vs. production of palatalized forms before *-i*.

	b	se(b)	z	p
(Intercept)	−1.3751	0.4113	−3.344	0.000827***
Voiceless	0.218	0.2077	1.049	0.294075
Labial Training	4.0657	0.7809	5.207	1.92E-07***
Judgment Test	−0.3777	0.4064	−0.929	0.352711
Labial Training × Judgment Test	−3.6816	0.786	−4.684	2.81E-06***
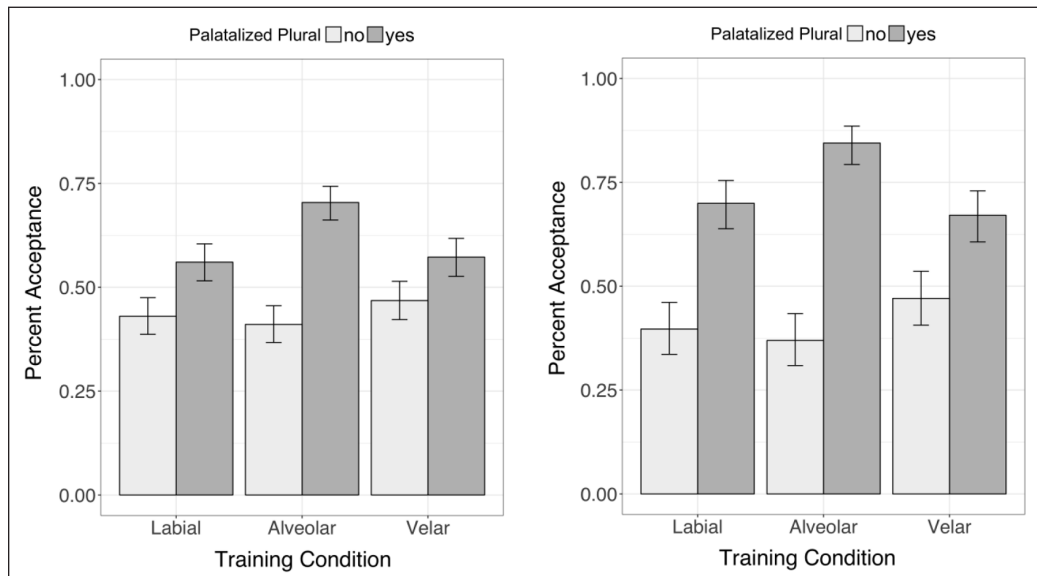
Figure 10: Judgments of correct palatalized plurals (dark bars) vs. incorrect faithful plurals (light bars) by training condition. Left panel: Across suffixes. Right panel: Before *-i*.

dissociation between production and judgment after exposure to labial palatalization (left bars): Labial palatalization is typically accepted but not produced (as shown by the significant interaction in **Table 7**,[12] which significantly improves the fit of the model, $\chi^2(1)$ = 21.94, p < 0.001). This dissociation between judgments and production is not present for training on alveolar and velar palatalization: When trained to palatalize alveolars or velars, participants produce palatalization as often as they accept it; when trained to palatalize labials, they accept palatalization but do not produce it.

One might argue that judgments are simply more lenient than production, hence speakers accept more than what they would produce (e.g., Kempen & Harbusch, 2005). However, a comparison of incorrectly faithful to-be-palatalized and correctly palatalized plurals, as shown in **Figure 10** below, shows that this explanation is insufficient to account for the present data. Judgments of faithful p→pi mappings are significantly lower than those of the unfaithful p→tʃi mappings following training on labial palatalization (left bars; $\chi^2(1)$ = 11.86, p < 0.001 before *-i*; $\chi^2(1)$ = 4.53, p = 0.03 across vowels). Labial palatalization is *preferred* to non-palatalization after training, and yet seldom produced. Judgment is not simply more lenient than production after labial palatalization training: The mapping preferred in judgment is dispreferred in production (see also Zuraw, 2000).

4. General Discussion

The full pattern of results is summarized in **Figures 11–13**. The production data in **Figure 11** show that the productivity of palatalization with to-be-palatalized consonants at test is highest when these consonants are alveolar, lower when they are velar, and lowest when they are labial. This difference suggests that larger changes are more difficult to produce, or learn to produce, than small changes (Hypothesis 2).

In contrast, production data provide little support for Hypothesis 3, the proposal that large changes are more likely to be overgeneralized. The palatalization rates of any given not-to-be-palatalized consonant are not strongly affected by the identities of the to-be-palatalized consonants. If anything, the overgeneralization patterns provide support

---

[12] dv ~ Test Voice + Labial Training * Test + (1 + Test Voice + Test|Subject) + (1|Base), restricted to ToBePalatalizedPlace == "yes", production *or* judgments of palatalized forms before *-i*.
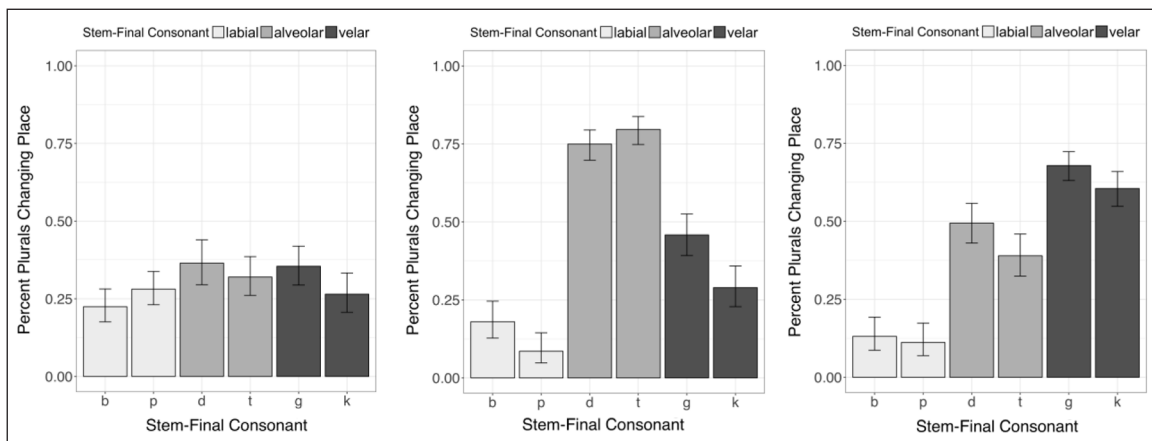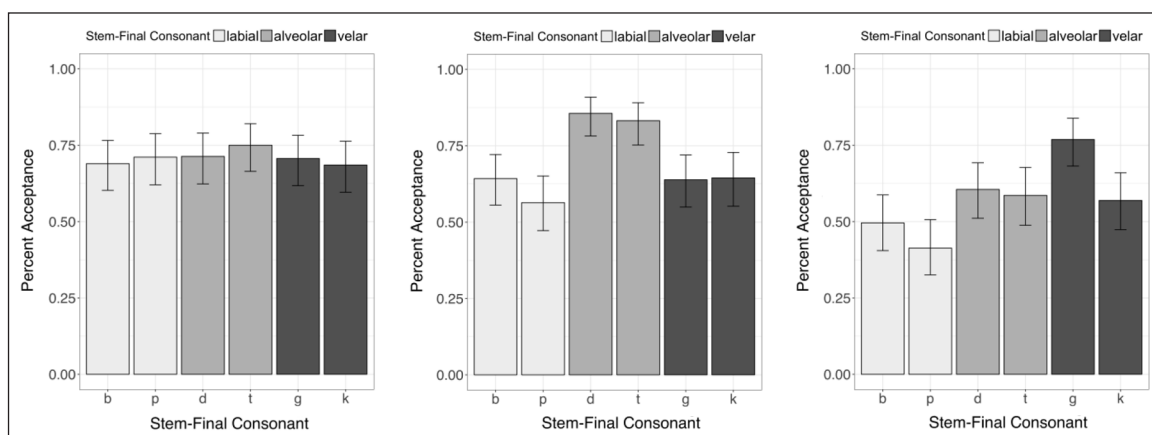
Figure 11: Differences in palatalization rates before *-i* in production across individual stops and training conditions; Labial Training is on the left, Alveolar in the center, and Velar on the right. Shading indicates place of articulation from labial (lightest) through alveolar (medium) to velar (darkest). Voiced consonants are on the left within shading, while voiceless ones are on the right.
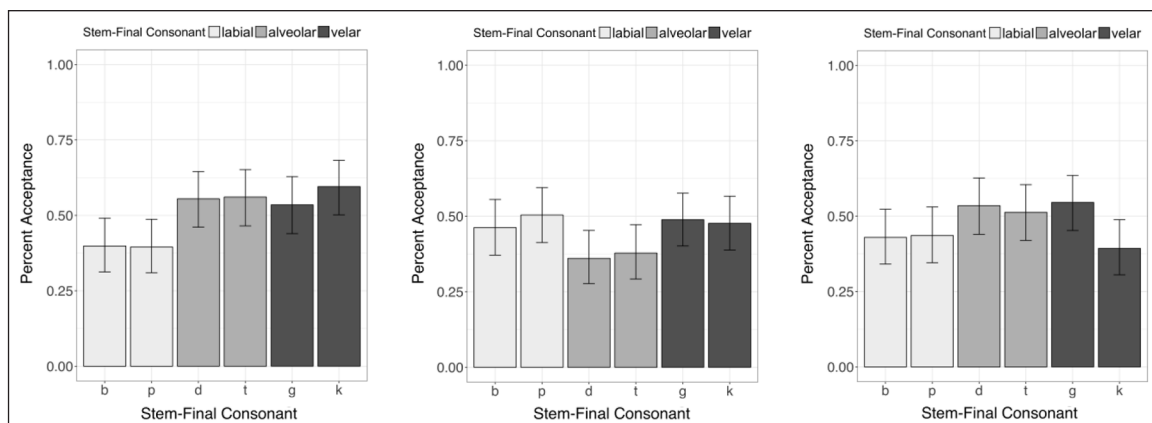
Figure 12: Differences in judgments of faithful mappings before *-i* across individual stops and training conditions; Labial Training is on the left, Alveolar in the center, and Velar on the right. Shading indicates place of articulation from labial (lightest) through alveolar (medium) to velar (darkest). Voiced consonants are on the left within shading, while voiceless ones are on the right.

Figure 13: Differences in judgments of palatalized plurals before *-i* across individual stops and training conditions; Labial Training is on the left, Alveolar in the center, and Velar on the right. Shading indicates place of articulation from labial (lightest) through alveolar (medium) to velar (darkest). Voiced consonants are on the left within shading, while voiceless ones are on the right.

for overgeneralization of a change to similar input sounds, whether or not those sounds are farther from or closer to the output of the change: In **Figure 6**, not-to-be-palatalized linguals are palatalized slightly more after training on lingual palatalization than after training on labial palatalization. In typology, Mielke (2008) has argued that alternations can spread from segment to segment by this kind of process of analogical change, resulting in phonologically active classes that cannot be described by a conjunction of distinctive features but instead have a 'family resemblance' structure. In our experiment, exposure to lingual palatalization patterns may increase the palatalization rates of similar lingual consonants more than does exposure to labial palatalization. Note that this interpretation may also apply to the overgeneralizations previously observed by White (2013, 2014). For example, in his studies, p→v overgeneralized to f→v more than b→v did. While this result is consistent with overgeneralization of saltatory alternations to intermediate sounds, it may also arise from the fact that [p] shares voicelessness with [f], whereas [b] does not.

Together, these results suggest that saltatory alternations are likely to be diachronically unstable because they lose productivity, not because they become overgeneralized to the 'jumped over' sounds. The English [k]/[s] in *electri*[k]/*electri*[s]*ity* is a case in point, showing limited productivity (Pierrehumbert, 2006) rather than being overgeneralized to the intermediate [t]. Thus, we expect that labial palatalization in languages that have it (e.g., Southern Bantu; Ohala, 1978) is likely to be lost through underapplication, disappearing from the language, rather than be generalized, resulting in indiscriminate palatalization of all consonants. Indeed, in a first study on the productivity of labial palatalization, it has recently been shown to be only partially productive for speakers of Xhosa (Bennett & Braver, 2015). For this reason, we believe that the difficulty of producing or learning to produce large changes is partially responsible for the typological rarity of large stem changes.

Figure 12 summarizes the results for judgments of faithful mappings like p→pi or t→ti, whereas **Figure 13** summarizes the results for judgments of changes like p→tʃi and t→tʃi. Training condition did not significantly affect the judgments of faithful mappings but did affect the judgments of changes, suggesting that differences between conditions are due to differences among changes and not the faithful outputs those changes avoid (Hypothesis 1). In Optimality-Theoretic terms, the dislike of p→tʃi is not due to the low markedness of [pi] and the liking of t→tʃi is not due to the high markedness of [ti]. Large changes are hard to learn and/or execute because they are large changes, not because they mutilate perfectly acceptable structures.

This conclusion is supported by our previous results reported in Stave et al. (2013). Whereas one could argue that sequences like [ki] and [ti] are marked compared to [tʃi], on either perceptual or articulatory grounds (Wilson, 2006), it is difficult to make the same argument for [ka] and [ta] vs. [tʃa]. Palatalization before -*a* is phonetically unmotivated and cannot be described as repairing a marked structure. Stave et al. (2013) examined palatalization in the pre-*a* environment. As in the present study, participants in that experiment found alveolar palatalization easiest to learn, and labial palatalization hardest to learn, with velar palatalization in between. Also, as in the present study, judgments of faithful mappings were equal across conditions, while judgments of changes varied.

The present study extends Stave et al.'s (2013) results by showing that even when palatalization can be seen as improving markedness, learnability differences between different kinds of palatalization cannot be described by markedness differences among the structures they repair. These results suggest that learnability of a novel alternation may not be strongly determined by whether that alternation repairs a phonotactic violation (cf. Prince & Smolensky, 1993/2004; Wilson, 2006). Indeed, previous studies that have looked for the link between phonotactics and alternations have often been unsuccessful

Smolek and Kapatsinski: What happens to large changes? Saltation
produces well-liked outputs that are hard to generate

Art. 10, page 19 of 27

in finding a significant difference (Chong, 2016; Pater & Tessier, 2003). Furthermore, in actual languages, the same alternation may be more productive in contexts where it is phonotactically unmotivated than in contexts where it is. For example, Kapatsinski (2010) shows that the k→tʃ alternation in Russian is more productive before -*ok* than before -*ik* or -*i*-. Even if the markedness of a structure avoided by an alternation may have some influence on its learnability (Mitrović, 2012; Wilson, 2006), learnability of an alternation does not reduce to the markedness of the structure it repairs. Uncontroversially, it is also affected by the statistical strengths of competing patterns in the lexicon (Kapatsinski, 2010). Our results suggest that it is also affected by the alternation's articulatory or perceptual magnitude (see also Gnanadesikan, 1997; Hayes & White, 2015; Kirchner, 1996; Skoruppa et al., 2011; Steriade, 2001/2009; White, 2013, 2014, 2017).

Judgments of alternations in **Figure 13** show some agreement with and some divergences from the production data in **Figure 11**. The data are in agreement that the magnitude of the experienced alternation influences mainly the to-be-palatalized consonants rather than the others; i.e., the experienced alternations are affected more than overgeneralizations. The judgment data therefore support the proposal that large changes are especially hard for a language to maintain because they are especially likely to lose productivity, and not because they are especially likely to be overgeneralized.

However, in another way, the judgments and production data are in disagreement. After exposure to labial palatalization, participants judge palatalization to be better than faithful retention of the input consonant, regardless of the consonant's identity. However, they are far more likely to faithfully retain the input than to palatalize it. In other words, judgments suggest that participants would like to palatalize everything, but production data shows that they can palatalize nothing. This is not an isolated anomaly. For example, Zuraw (2000) found that Tagalog speakers prefer novel prefixed words that have undergone nasal substitution, a stem change, to those that have not, but tend not to produce nasal substitution when generating such words themselves. White (2013) observed that accounting for his production data required introducing a *Alternate constraint that was not needed to describe judgments. These kinds of dissociations between production and judgment appear to introduce some uncertainty regarding the fate of large changes. While these changes are likely to be leveled in production, the resulting faithful forms may then be judged unacceptable by the listener. If the speaker heeds the listener's judgment, they may then avoid the faithful output in future productions. Conversely, an over-extension of palatalization to a new input consonant may be preferred by the listener over a faithful output, with the listener's judgment rewarding the speaker for over-extending palatalization.

Several recent studies have suggested that speakers do adjust their productions in response to listener feedback (Buz, Tanenhaus, & Jaeger, 2016; Goldstein, King, & West, 2003; Maniwa, Jongman, & Wade, 2009; Schertz, 2013; Seyfarth, Buz, & Jaeger, 2016; Warlaumont, Richards, Gilkerson, & Oller, 2014). Indeed, White (2013) has used listener feedback to train learners to produce alternations. These results suggest that listeners' beliefs about speakers' productions, if made apparent to and heeded by the speaker, can influence the speakers' future productions. On the other hand, listeners' beliefs about what is and is not acceptable are also based on the productions they experience, so that it is often the case that, in language change, "use leads, and belief follows" (Harmon & Kapatsinski, 2017). The sociolinguistic literature is full of dissociations between judgment and production, so that speakers who routinely produce an innovative form nonetheless judge it to be unacceptable due to social stigma associated with it (Labov, 1996). Yet, there is little evidence on whether these judgments, despite being internalized by the speakers, result in avoidance of the forms judged unacceptable or have the power to limit their spread (see Curzan, 2014, for a review).

More research on the interaction between use and belief is sorely needed. In addition to careful observational studies of the impact of social acceptability on use, experimental work should investigate how judgment and production interact by examining more interactive tasks (e.g., Buz et al., 2016) and/or varying the order of production and judgment tasks (Harmon & Kapatsinski, 2017), and research on language acquisition outside of the laboratory should examine the timecourse of development of judgment and production in the acquisition of alternations (e.g., Kerkhoff, 2007).

5. Limitations and future directions

The present study is not without its limitations. These limitations stem from the fact that our participants are not blank slates but American English speakers. They may well generalize from their knowledge of English to the miniature artificial language they are exposed to and, perhaps, even impose English patterns on the language (e.g., Finn & Hudson Kam, 2008). Using native English speakers as the study population allows us to compare results to perceptual data from Guion (1998) and to previous results on palatalization learning obtained by Wilson (2006), Kapatsinski (2012, 2013), and Stave et al. (2013). However, it also leaves the observed biases susceptible to explanations based on first language phonological experience rather than differences in change magnitudes (see also Skoruppa et al., 2011, for similar concerns regarding their alternations; though cf. Garcia, van Horne, & Hartshorne, 2017; Mitrović, 2012; Wang & Saffran, 2014; for evidence against first-language transfer in miniature artificial language learning). In particular, English has alveolar palatalization patterns that are productive in specific contexts: before glides in frequent phrases like *would you* and *bet you* and in words like *creature* (cf. *create*) or *torture* (cf. *extort*). While the former do not involve a complete change in place of articulation (Zsiga, 1995), and the latter are of doubtful productivity, the existence of such patterns may have made alveolar palatalization easier to learn for our subjects. It would be interesting to investigate whether alveolar palatalization is also easier to learn in a miniature artificial language for speakers whose native languages lack alveolar palatalization altogether.

The data also provide evidence, albeit somewhat limited, that palatalization of [g] was favored over palatalization of [k]. The difference reaches significance in the middle panel of **Figure 11** and the right panel of **Figure 12**, and has also been observed by Wilson (2006). This asymmetry cannot be explained by perceptual change magnitude because [ki] and [tʃi] are more perceptually confusable than [gi] and [dʒi] for English speakers (Guion, 1998; see Wilson, 2006, for discussion). It may instead be due to a first-language influence, namely the influence of English spelling-sound correspondence patterns. In English, orthographic <g> often maps onto [dʒ] (~30% of <g>'s are [dʒ] and ~70% are [g], Gontijo, P. F., Gontijo, I., & Shillcock, 2003). In contrast, <k> always maps onto [k], <c> maps onto a palatal only 3% of the time, and <ch> maps onto [tʃ] 87% of the time. Thus spellings of [k] and [tʃ] are largely distinct, while spellings of [g] and [dʒ] are often the same, which may lead literate English speakers to categorize the latter together. According to Moreton & Pater (2012a), it is easier to learn alternations between sounds grouped into a single category, which may account for the otherwise unexpected asymmetry between [k] and [g]. It would be interesting to see whether the difference between /k/ and /g/ observed here and in Wilson (2006) is nullified or reversed with speakers of other languages or preliterate children.

Velar palatalization may perhaps be thought to be favored over labial palatalization because /k/ can change into [s] in English, as in *electric-electricity*, while [p] and [b] never participate in any alternations. However, we do not consider it likely that this is

Smolek and Kapatsinski: What happens to large changes? Saltation
produces well-liked outputs that are hard to generate

Art. 10, page 21 of 27

responsible for the aversion to changing labials we observed. The process is productive in only very restricted circumstances, namely when the input is Latinate-sounding and a specific Latinate suffix is attached. The stimuli used here and in Stave et al. (2013) bear little resemblance to the words that exemplify k→s alternations in English. With stimuli like ours, English speakers are very reluctant to change /k/ into [s] even before change-triggering suffixes like –ity (Pierrehumbert, 2006). Nonetheless, future work should examine acquisition of patterns absent from English altogether and to study palatalization by participants whose native language has no productive palatalization process and no tendency for alternations to involve non-labials.

Let us imagine for a moment that future cross-linguistic work shows that labial palatalization is hard to learn for English speakers because they know labials to be relatively unchangeable. We would then conclude that first language experience endows English speakers with knowledge that favors some alternations over others. Our results on the judgments of faithful mappings indicate that this knowledge does not reduce to knowledge of phonotactics or product-oriented schemas, i.e., knowledge about what sounds and sound combinations are more or less common in the language (cf. Bybee, 2001). At a minimum, the present data suggest that English speakers must know that labials are less changeable than velars, which are less changeable than alveolars. However, White's (2013) data show no evidence for alternations involving labial inputs (e.g., p→v) to be harder to learn than alternations involving alveolar inputs (e.g., tàð). In combination, then, these studies may suggest that learners assign prior probabilities to alternations, or paradigmatic mappings, as suggested by the *Map constraints of Zuraw (2007) or the operations of rule-based phonology (e.g., Chomsky & Halle, 1965; Labov, 1969).

Previous work has suggested that palatalization is harder to learn when it is phonetically unmotivated in context (Mitrović, 2012; Wilson, 2006). However, it is not yet clear whether this influence of context segments is independent of the identity of the input segment. Do speakers assign probabilities to rules (changes in context such as p→tʃ/_a) or do they assign probabilities to changes (p→tʃ) and their outputs (tʃa) and then combine them to evaluate the probability of a particular change resulting in a particular output? The comparison between results reported here and in Stave et al. (2013) suggests that the prior probability of an alternation is at least partially context-independent: p→tʃ is harder to learn than k→tʃ, which is harder than t→tʃ, whether the palatalization is triggered by -i or -a. However, this comparison is less straightforward than we would like because the languages used in Stave et al. (2013) and here differ in more than the identity of the palatalizing suffix. An interesting direction for future work is to expose participants to languages in which the magnitude of a change and the context in which the change occurs are the only factors manipulated.

Whether or not the biases observed here can be explained by first native language experience, one interesting difference between these results and those obtained by White (2013, 2014) is that the biases observed here manifest themselves largely in no-change errors on the segments that the participants are trained to change. Likewise, the differences between the large-change and small-change conditions in Skoruppa et al. (2011) came from errors on the trained consonants, most of which were no-change errors. In contrast, White's (2013, 2014) participants displayed their biases in patterns of overgeneralization: Large changes were overgeneralized more than small changes. We suggest that this discrepancy is due to exclusion of participants who made too many no-change errors on the trained segments in White's studies. These exclusions tended to affect the large-change condition more than the small-change condition (e.g., White, 2013, p. 72), suggesting that including these participants would have resulted in a larger proportion of no-change

errors after exposure to large changes. However, if the results of our study and those of Skoruppa et al. (2011) are explainable by first language experience, whereas those of White (2013, 2014) are not, then it may be possible that familiarity and magnitude affect the learnability of a change and the likelihood of overgeneralization respectively. Unfamiliar changes are harder to notice in training data and to execute in production, whereas large changes are more likely to be overgeneralized than small changes.

We consider this explanation for the differences between the studies to be unlikely for two reasons. First, it is not clear that the alternations presented to White's (2013, 2014) participants are entirely novel, either. White compared changes involving turning a voiceless stop into a voiced fricative between vowels to intervocalic lenition of a voiced stop. While English does not have intervocalic stop lenition as a categorical process, variable lenition is quite common (e.g., Davidson, 2011; Honeybone, 2001; Riebold, 2011; Sangster, 2001; Warner & Tucker, 2011) and UCLA students in White's study are likely to have some exposure to Spanish, which does have voiced stop lenition, and Spanish-accented English (e.g., Zampini, 1996). These lenition processes tend to preserve voicing, which makes the smaller changes potentially more familiar than the larger changes. Second, as discussed earlier, diachronic data suggest that large changes are likely to lose productivity rather than be overgeneralized to new sounds. Nonetheless, it appears important to replicate the present results with speakers of different languages. Of particular interest would be speakers of languages with labial palatalization (e.g., Southern Bantu; Braver & Bennett, 2015; Ohala, 1978). Mitrović (2012) found that speakers of Serbian, whose native language has productive velar palatalization before [e] but not [i], nonetheless show a learning bias for palatalization before [i] in a miniature artificial language, a finding that provides particularly strong support for this learning bias. It would be interesting to likewise pit familiarity and change magnitude against each other. Another population of interest would be speakers for whom none of the alternative palatalization patterns are productive.

6. Conclusion

Languages don't seem to like large alternations skipping over sounds, like the k~s alternation of *electric/electricity*, which skips over [t]. Previous research has hypothesized that the typological rarity of such saltatory alternations is due in part to their diachronic instability. In particular, saltatory patterns may be especially likely to be overgeneralized, spreading to intermediate sounds. For example, k~s is in danger of spawning t~s, which would make the pattern non-saltatory. This proposed diachronic trajectory is supported by findings that saltatory alternations are likely to be overgeneralized to intermediate sounds. However, previous studies reporting this result have either trained learners to criterion on the trained alternation (ensuring they would learn k~s as well as its non-saltatory counterpart), or excluded participants who failed to reach a criterion from analysis. When all participants are included, the results suggest an alternative diachronic trajectory. The outcomes of saltatory alternations are judged to be as acceptable as those of non-saltatory alternations, and both are more acceptable than their faithful competitors. However, an unfaithful outcome is less likely to be generated when it is the outcome of a saltatory alternation. If English speakers learned a t~s alternation, and *electricity* were formed from *electrite*, it would be just as good (and just as superior to *electrickity*) as it is now. However, it would be much easier to generate. Instead of being overgeneralized, saltatory alternations are likely to lose productivity faster than non-saltatory alternations, disappearing from the language. Large changes are hard to make. Those who experience such changes may think you should change everything, but at the same time are likely to change nothing. The outcome of a smaller change may be less preferred, but such changes are easier to produce and to keep producing.

Smolek and Kapatsinski: What happens to large changes? Saltation produces well-liked outputs that are hard to generate

Art. 10, page 23 of 27

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** PDF file with training stimuli. DOI: https://doi.org/10.5334/labphon.93.s1
- **Appendix B.** PDF file with analyses using untransformed data. DOI: https://doi.org/10.5334/labphon.93.s2
- **Data.** CSV file with full dataset. DOI: https://doi.org/10.5334/labphon.93.s3
- **Code.** R file with analysis code. DOI: https://doi.org/10.5334/labphon.93.s4

Competing Interests

The authors have no competing interests to declare.

References

Andersen, H. 1973. Abductive and deductive change. *Language, 49*(4), 765–793. DOI: https://doi.org/10.2307/412063

Bateman, N. 2007. *A crosslinguistic investigation of palatalization* (Doctoral dissertation, UCSD).

Bates, D., Maechler, M., Bolker, B., & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Bennett, W. G., & Braver, A. 2015. Phonology or morphology: Inter-speaker differences in Xhosa labial palatalization. *Paper presented at the Annual Meeting of the Linguistic Society of America.* Portland, OR.

Bhat, D. N. 1978. A general study of palatalization. *Universals of human language, 2,* 47–92.

Bolognesi, R. 1998. *The phonology of Campidanian Sardinian: A unitary account of a self-organizing structure.* The Hague: Holland Institute for Generative Linguistics.

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. 2016. Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language, 89,* 68–86. DOI: https://doi.org/10.1016/j.jml.2015.12.009

Bybee, J. 2001. *Phonology and language use.* Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511612886

Bybee, J. 2008. Formal universals as emergent phenomena: The origins of structure preservation. In: Good, J. (ed.), *Linguistic universals and language change*, 108–121. Oxford: Oxford University Press.

Bybee, J. 2010. *Language, usage and cognition.* Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511750526

Bybee, J., & Slobin, D. I. 1982. Why small children cannot change language on their own: Suggestions from the English past tense. In: Alqvist, A. (ed.), *Papers from the 5th International Conference on Historical Linguistics*, 29–37. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/cilt.21.07byb

Chen, M. 1973. Predictive power in phonological description. *Lingua, 32,* 173–191. DOI: https://doi.org/10.1016/0024-3841(73)90041-7

Chomsky, N., & Halle, M. 1965. Some controversial questions in phonological theory. *Journal of Linguistics, 1*(2), 97–138. DOI: https://doi.org/10.1017/S0022226700001134

Chong, A. J. 2016. Learning consequences of derived environment effects. *Paper presented at the 90th Annual Meeting of the Linguistic Society of America.* Washington, DC, January 7. DOI: https://doi.org/10.3765/plsa.v1i0.3709

Crosswhite, K. M. 2000. Vowel reduction in Russian: A unified account of standard, dialectal, and "dissimilative" patterns. In: Crosswhite, K. M., & McDonough, J. (eds.),

University of Rochester Working Papers in the Language Sciences, 1, 107–172. Spring. Rochester, NY: University of Rochester.

Curzan, A. 2014. *Fixing English: Prescriptivism and language history.* Cambridge: Cambridge University Press.

Davidson, L. 2011. Characteristics of stop releases in American English spontaneous speech. *Speech Communication, 53*(8), 1042–1058. DOI: https://doi.org/10.1016/j.specom.2011.05.010

Do, Y. A. 2013. *Biased learning of phonological alternations.* (Doctoral Dissertation, MIT).

Finley, S. 2008. *Formal and cognitive restrictions on vowel harmony.* (Doctoral Dissertation, Johns Hopkins University).

Finn, A. S., & Hudson Kam, C. L. 2008. The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition, 108*(2), 477–499. DOI: https://doi.org/10.1016/j.cognition.2008.04.002

Garcia, C., van Horne, K. D., & Hartshorne, J. 2017. Replication of Finn & Hudson Kam (2008) The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation, Exp. 1. psyarxiv. DOI: https://doi.org/10.17605/OSF.IO/2XCWK

Gnanadesikan, A. E. 1997. *Phonology with ternary scales* (Doctoral dissertation, University of Massachusetts, Amherst).

Goldstein, M. H., King, A. P., & West, M. J. 2003. Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences, 100*(13), 8030–8035. DOI: https://doi.org/10.1073/pnas.1332441100

Gontijo, P. F., Gontijo, I., & Shillcock, R. 2003. Grapheme—phoneme probabilities in British English. *Behavior Research Methods, Instruments, & Computers, 35*(1), 136–157. DOI: https://doi.org/10.3758/BF03195506

Guion, S. G. 1998. The role of perception in the sound change of velar palatalization. *Phonetica, 55*(1–2), 18–52. DOI: https://doi.org/10.1159/000028423

Harmon, Z., & Kapatsinski, V. 2017. Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology, 98,* 22–44. DOI: https://doi.org/10.1016/j.cogpsych.2017.08.002

Hayes, B., & White, J. 2015. Saltation and the P-map. *Phonology, 32*(2), 1–36. DOI: https://doi.org/10.1017/S0952675715000159

Honeybone, P. 2001. Lenition inhibition in Liverpool English. *English Language & Linguistics, 5*(2), 213–249. DOI: https://doi.org/10.1017/S1360674301000223

Hudson Kam, C. L., & Newport, E. L. 2009. Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*(1), 30–66. DOI: https://doi.org/10.1016/j.cogpsych.2009.01.001

Kapatsinski, V. 2007. To scheme or to rule: Evidence against the Dual-Mechanism Model. In: *Proceedings of the 31st Annual Meeting of the Berkeley Linguistics Society,* 193–204.

Kapatsinski, V. 2010. Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology, 1*(2), 361–393. DOI: https://doi.org/10.1515/labphon.2010.019

Kapatsinski, V. 2012. What statistics do learners track? Rules, constraints and schemas in (artificial) grammar learning. In: Th Gries, S., & Divjak, D. (eds.), *Frequency effects in language learning and processing,* 53–73. Berlin: Mouton de Gruyter.

Kapatsinski, V. 2013. Conspiring to mean: Experimental and computational evidence for a usage-based harmonic approach to morphophonology. *Language, 89*(1), 110–148. DOI: https://doi.org/10.1353/lan.2013.0003

Kempen, G., & Harbusch, K. 2005. The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German

Smolek and Kapatsinski: What happens to large changes? Saltation produces well-liked outputs that are hard to generate

Art. 10, page 25 of 27

clauses. In: Pechmann, T., & Habel, C. (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives,* 329–349. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110197549.329

Kerkhoff, A. O. 2007. *Acquisition of morpho-phonology: The Dutch voicing alternation* (Doctoral dissertation, University of Nijmegen).

Kirchner, R. 1996. Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry*, *27*(2), 341–350.

Kochetov, A. 2011. Palatalisation. In: Ewen, C., Hume, B., van Oostendorp, M., & Rice, K. (eds.), *Blackwell Companion to Phonology*, 1666–1690. Wiley-Blackwell.

Krajewski, G., Theakston, A. L., Lieven, E. V., & Tomasello, M. 2011. How Polish children switch from one case to another when using novel nouns: Challenges for models of inflectional morphology. *Language and Cognitive Processes*, *26*(4–6), 830–861. DOI: https://doi.org/10.1080/01690965.2010.506062

Labov, W. 1969. Contraction, deletion, and inherent variability of the English copula. *Language*, *45*(4), 715–762. DOI: https://doi.org/10.2307/412333

Labov, W. 1996. When intuitions fail. *Chicago Linguistic Society*, *32*, 76–106.

Luce, P. A., & Pisoni, D. B. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36. DOI: https://doi.org/10.1097/00003446-199802000-00001

Maniwa, K., Jongman, A., & Wade, T. 2009. Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America, 125*(6), 3962–3973. DOI: https://doi.org/10.1121/1.2990715

Mielke, J. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Mitrović, I. 2012. A phonetically natural vs. native language pattern: An experimental study of velar palatalization in Serbian. *Journal of Slavic Linguistics*, *20*(2), 229–268. DOI: https://doi.org/10.1353/jsl.2012.0011

Moreton, E., & Pater, J. 2012a. Structure and Substance in Artificial-phonology Learning, Part I: Structure. *Language and Linguistics Compass*, *6*(11), 686–701. DOI: https://doi.org/10.1002/lnc3.363

Moreton, E., & Pater, J. 2012b. Structure and substance in artificial-phonology learning, Part II: Substance. *Language and Linguistics Compass*, *6*(11), 702–718. DOI: https://doi.org/10.1002/lnc3.366

Ohala, J. 1978. Southern Bantu vs. the world: The case of palatalization of labials. *Berkeley Linguistics Society*, *4*, 370–386. DOI: https://doi.org/10.3765/bls.v4i0.2218

Pater, J., & Tessier, A. M. 2003. Phonotactic knowledge and the acquisition of alternations. *Proceedings of the 15th International Congress on Phonetic Sciences*, 1177–1180.

Pater, J., & Tessier, A. M. 2006. L1 phonotactic knowledge and the L2 acquisition of alternations. In: Slabakova, R., Montrul, S., & Prévost, P. (eds.), *Inquiries in linguistic development: Studies in honor of Lydia White,* 115–131. John Benjamins. DOI: https://doi.org/10.1075/z.133.07pat

Peperkamp, S., Le Calvez, R., Nadal, J. P., & Dupoux, E. 2006. The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition, 101*(3), B31–B41. DOI: https://doi.org/10.1016/j.cognition.2005.10.006

Pierrehumbert, J. B. 2006. The statistical basis of an unnatural alternation. In: Goldstein, L., Whalen, D. H., & Best, C. (eds.), *Laboratory phonology, 8,* 81–107. Mouton de Gruyter.

Prince, A., & Smolensky, P. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Wiley.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Art. 10, page 26 of 27

Smolek and Kapatsinski: What happens to large changes? Saltation produces well-liked outputs that are hard to generate

Riebold, J. M. 2011. Time to pull out the stops: Spirantization in Pacific Northwestern English. *The Journal of the Acoustical Society of America*, *129*(4), 2453–2453. DOI: https://doi.org/10.1121/1.3588057

Sangster, C. M. 2001. Lenition of alveolar stops in Liverpool English. *Journal of Sociolinguistics*, *5*(3), 401–412. DOI: https://doi.org/10.1111/1467-9481.00156

Schane, S. A., Tranel, B., & Lane, H. 1975. On the psychological reality of a natural rule of syllable structure. *Cognition*, *3*(4), 351–358. DOI: https://doi.org/10.1016/0010-0277(74)90004-3

Schertz, J. 2013. Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, *41*(3), 249–263. DOI: https://doi.org/10.1016/j.wocn.2013.03.007

Seyfarth, S., Buz, E., & Jaeger, T. F. 2016. Dynamic hyperarticulation of coda voicing contrasts. *The Journal of the Acoustical Society of America*, *139*(2), EL31–EL37. DOI: https://doi.org/10.1121/1.4942544

Skoruppa, K., Lambrechts, A., & Peperkamp, S. 2011. The role of phonetic distance in the acquisition of phonological alternations. In: *Proceedings of NELS, 39*, 464–475.

Stave, M., Smolek, A., & Kapatsinski, V. 2013. Inductive bias against stem changes as perseveration: Experimental evidence for an articulatory approach to output-output faithfulness. In: *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*, 3454–3459.

Steriade, D. 2001/2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In: Hanson, K., & Inkelas, S. (eds.), *The nature of the word: Studies in honor of Paul Kiparsky*, 151–170. Cambridge, MA: MIT Press.

Tomas, E., van de Vijver, R., Demuth, K., & Petocz, P. 2017. Acquisition of nominal morphophonological alternations in Russian. *First Language*, *37*(5), 453–474. DOI: https://doi.org/10.1177/0142723717698839

Wagenmakers, E. J. 2007. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. DOI: https://doi.org/10.3758/BF03194105

Wang, T., & Saffran, J. R. 2014. Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, *5*. DOI: https://doi.org/10.3389/fpsyg.2014.00953

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. 2014. A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*(7), 1314–1324. DOI: https://doi.org/10.1177/0956797614531023

Warner, N., & Tucker, B. V. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, *130*, 1606. DOI: https://doi.org/10.1121/1.3621306

White, J. 2013. *Bias in phonological learning: Evidence from saltation* (Doctoral Dissertation, UCLA).

White, J. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition*, *130*(1), 96–115. DOI: https://doi.org/10.1016/j.cognition.2013.09.008

White, J. 2017. Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language*. DOI: https://doi.org/10.1353/lan.2017.0001

White, J., & Sundara, M. 2014. Biased generalization of newly learned phonological alternations by 12-month-old infants. *Cognition*, *133*(1), 85–90. DOI: https://doi.org/10.1016/j.cognition.2014.05.020

Wilson, C. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945–982. DOI: https://doi.org/10.1207/s15516709cog0000_89

Smolek and Kapatsinski: What happens to large changes? Saltation produces well-liked outputs that are hard to generate

Art. 10, page 27 of 27

Yun, G. H. 2006. *The interaction between palatalization and coarticulation in Korean and English* (Doctoral dissertation, University of Arizona).

Zampini, L. M. 1996. Voiced stop spirantization in the ESL of native speakers of Spanish. *Applied Psycholinguistics, 17,* 335–354. DOI: https://doi.org/10.1017/S0142716400007979

Zsiga, E. C. 1995. An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In: Connell, B., & Arvaniti, A. (eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV,* 282–302. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511554315.020

Zuraw, K. 2000. *Patterned exceptions in phonology* (Doctoral Dissertation, UCLA).

Zuraw, K. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog. *Language, 83,* 277–316. DOI: https://doi.org/10.1353/lan.2007.0105